ตามไปดู ISED '88

งานประชุมวิชาการ อิเล็กทรอนิกส์ดิกชันนารี

กองบรรณาธิการ

เมื่อเร็ว ๆ นี้ผู้เขียนมีโอกาสไปร่วมงานประชุมทางวิชาการเรื่องอิเล็กทรอนิกส์ดิกชันนารี ซึ่งจัดขึ้นเป็นครั้งแรก ณ ประเทศญี่ปุ่น ผู้เขียนเห็นว่างานนี้เป็นเรื่องที่มีบทบาทและความสำคัญสูงมากในอนาคต จึงใคร่ขอนำรายละเอียดบางอย่างมาเล่าสู่กันฟัง

เนื่องจากความจำเป็นของการใช้ข้อมูลและข่าวสารในยุคปัจจุบัน ทำให้มีการค้นคิดและพัฒนางานด้านประมวลผลก้าวหน้าไปมาก ความต้องการที่สำคัญของยุคข้อมูลและข่าวสารก็คือการประมวลผลภาษาธรรมชาติ ปัจจุบันงานที่เกี่ยวกับภาษาธรรมชาติเริ่มมีบทบาทกว้างขวางยิ่งขึ้น และมีผลต่อการผลิตเพื่อการค้ากันอย่างมากมาย ตั้งแต่ระดับคำซึ่งกลายมาเป็นสินค้าแฝงในเวิร์ดโปรเซสเซอร์ การตรวจสอบตัวสะกด การหาคำพ้อง (thesaurus) การตรวจสอบไวยากรณ์ การถามตอบกับเครื่องจักรและทำให้เครื่องจักรฉลาดขึ้น ระบบการแปลภาษาด้วยเครื่อง ฯลฯ

บรรยากาศของห้องประชุมที่มีผู้สนใจจากทั่วโลกมากกว่า 40 ประเทศ

ในประเทศญี่ปุ่นมีบริษัทคอมพิวเตอร์หลายบริษัทที่เกี่ยวข้องกับการสร้างอิเล็กทรอนิกส์ดิกชันนารี โดยเฉพาะอย่างยิ่งทุกบริษัทมีโครงการพัฒนาซอฟต์แวร์ของตนเอง ดังนั้นจึงมีการพัฒนาระบบประมวลผลภาษาธรรมชาติและมีแนวโน้มของการลงทุนสูงขึ้นเป็นลำดับ รัฐบาลญี่ปุ่นโดยกระทรวงการค้าและอุตสาหกรรม (MITI) ได้เข้ามามีบทบาทโดยเป็นแกนลงทุนร่วมกับบริษัทคอมพิวเตอร์อีกหลายบริษัท แล้วตั้งบริษัทใหม่ขึ้นมาชื่อว่า Japan Electronic Dictionary Research Institute LTD หรือเรียกย่อ ๆ ว่า EDR การจัดประชุมสัมมนาครั้งนี้แกนกลางการจัดประชุมประกอบด้วย CICC และสถาบันที่เกี่ยวข้องกับการส่งเสริมการผลิตซอฟต์แวร์และข่าวสารข้อมูลอีก 8 สถาบัน ระหว่างวันที่ 24-25 พฤศจิกายน 2531 ซึ่งเป็นวันก่อนหน้าการจัดประชุมเรื่องคอมพิวเตอร์ยุคที่ 5 ส่วนสถานที่จัดอยู่แห่งเดียวกับงานนิทรรศการและการประชุมเรื่องคอมพิวเตอร์ยุคที่ 5 ณ โรงแรมปริ๊นซ์ ใจกลางกรุงโตเกียว

ผู้บรรยายที่ได้รับเชิญมาล้วนแล้วแต่เป็นมือหนึ่งของโลกทางด้านการประมวลผลภาษาธรรมชาติทั้งสิ้น โดยมี Makoto Nagao ประธานในที่ประชุมและเป็นผู้บรรยายนำ (keynote speech) ในหัวข้อความจำเป็นของอิเล็กทรอนิกส์ดิกชันนารี หลังจากนั้นช่วงเช้าของวันแรกเป็นการกล่าวถึงสถานภาพของการพัฒนาอิเล็กทรอนิกส์ดิกชันนารีในแต่ละประเทศ ซึ่งมีผู้บรรยายหลายคนจากประเทศญี่ปุ่นบรรยายโดย T.Yokoi ประเทศแคนาดาบรรยายโดย P.Isabell ประเทศอิตาลีบรรยายโดย A.Zampolli ประเทศในกลุ่ม ECC บรรยายโดย J.Geoscholekx ส่วนของเยอรมนีบรรยายโดย W.Lenders ประเทศจีนบรรยายโดย Dong อินโดนีเซียบรรยายโดย Darmawan มาเลเซียบรรยายโดย Zaki เดนมาร์กบรรยายโดย B.Macgood สำหรับประเทศไทยบรรยายโดย ดร. เพียรศิริ วงศ์วิภานนท์ ซึ่งการประชุมครั้งนี้มีคนไทยเพียงสองคนเท่านั้นที่เข้าร่วม

ในช่วงบ่ายวันแรกมีการบรรยายเรื่องการใช้อิเล็กทรอนิกส์ดิกชันนารีในการแปลภาษาด้วยเครื่องโดยเน้นระบบแปลภาษาด้วยเครื่องที่กำลังพัฒนากันอยู่ทั่วโลกได้แก่ ระบบ Mu ของมหาวิทยาลัยเกียวโต ATLAS ของบริษัทฟูจิตสึ AS-TRANSAC ของบริษัทโตชิบา PENSÉE ของบริษัทโอกิ HICATS ของบริษัทฮิตาชิ SYSTRAN ของบริษัท SYSTRAN และ ROSSETA ของบริษัทฟิลลิปส์

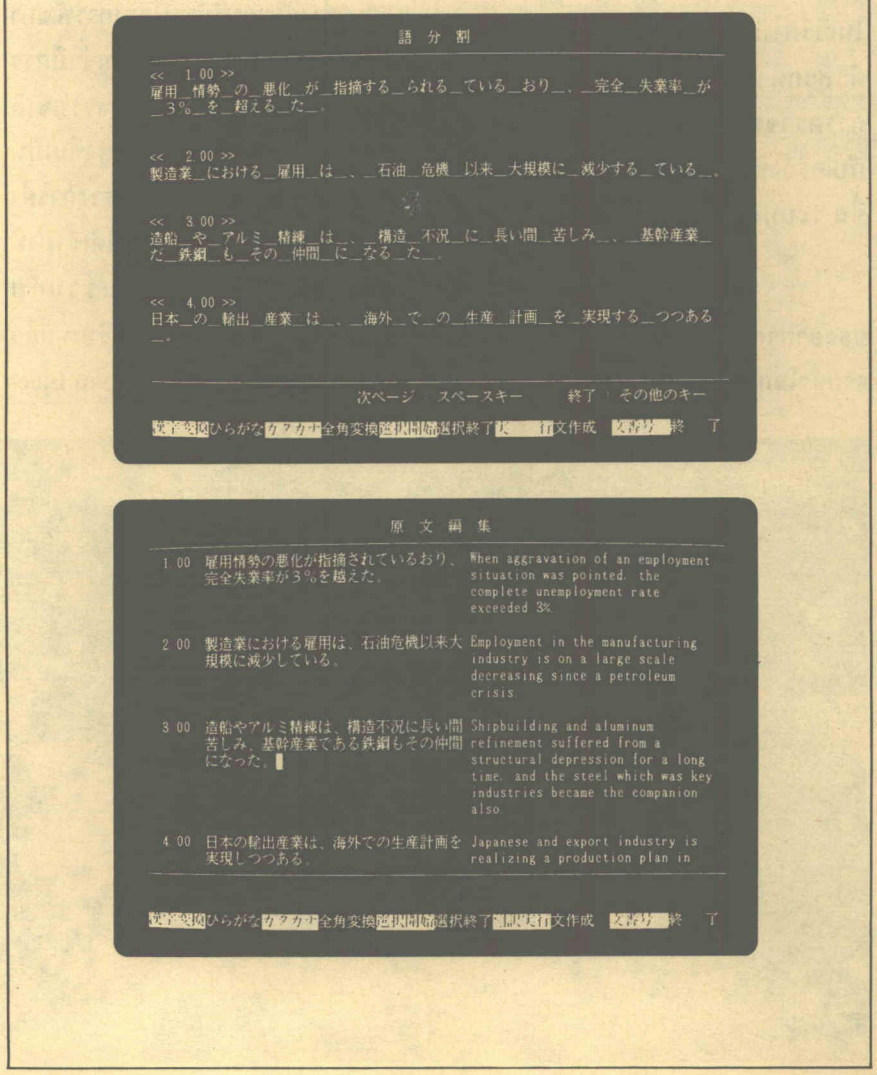
สำหรับการประชุมสัมมนาในวันที่สองเน้นหนักด้านการใช้ knowledge description ในงานอิเล็กทรอนิกส์ดิกชันนารี ผู้บรรยายคนสำคัญ เช่น Carbonell จากมหาวิทยาลัยคาร์เนกี้-เมลลอน และผู้บรรยายที่มีชื่อเสียงอีกหลายคน

ส่วนภาคบ่ายเป็นเรื่องบทบาทของอิเล็กทรอนิกส์ดิกชันนารีที่มีความสำคัญต่อสังคมมนุษย์ โดยมีการมองในมุมต่าง ๆ กันหลายด้าน นับว่าเป็นเรื่องน่าสนใจอยู่ไม่น้อย

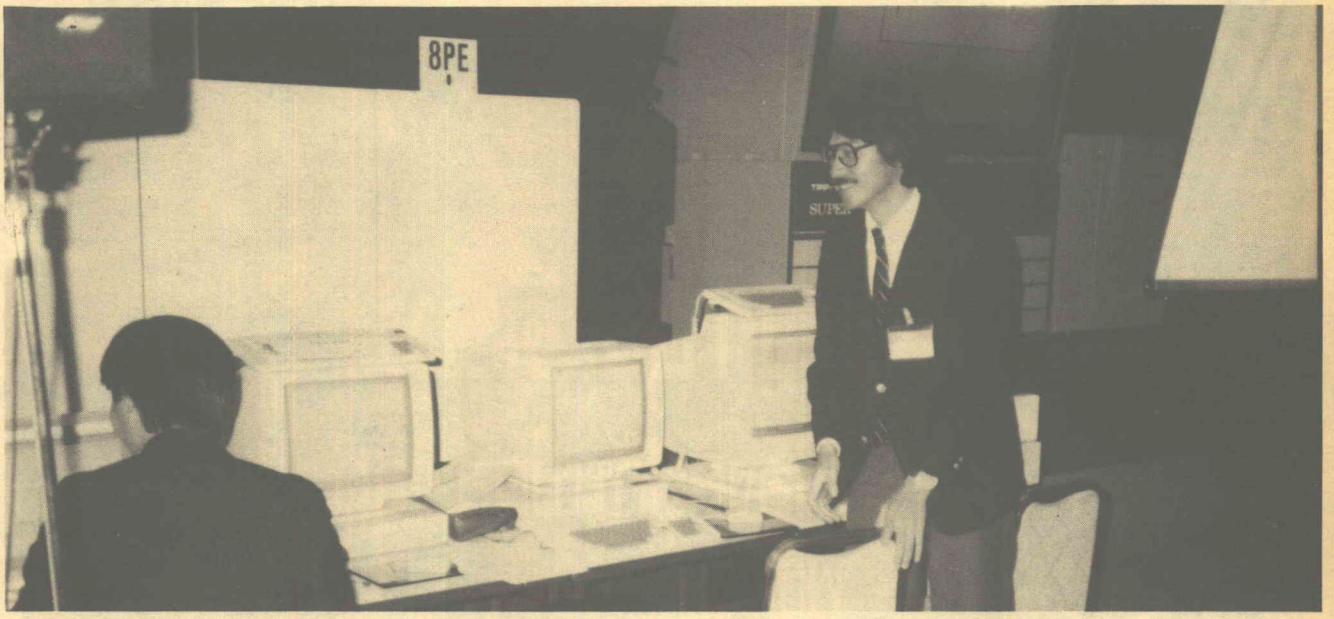
ผู้เขียนได้กล่าวถึงโครงร่างของงานแล้ว ในที่นี้ขออธิบายเพิ่มเติมถึงเรื่องอิเล็กทรอนิกส์ดิกชันนารีว่าเป็นอย่างไร โดยแยกอิเล็กทรอนิกส์ดิกชันนารีออกเป็นสองกลุ่มใหญ่คือ อิเล็กทรอนิกส์ดิกชันนารีสำหรับมนุษย์ใช้ และอิเล็กทรอนิกส์ดิกชันนารีสำหรับเครื่องจักรหรือคอมพิวเตอร์ใช้

อิเล็กทรอนิกส์ดิกชันนารีสำหรับให้มนุษย์ใช้ก็คือ ดิกชันนารีหรือพจนานุกรมที่เราใช้กันเสมอ โดยเก็บคำศัพท์และความหมายไว้ แต่เก็บรวมไว้ในสื่อทางคอมพิวเตอร์ เช่น พจนานุกรมฉบับราชบัณฑิตยสถานที่ผู้เขียนได้ร่วมกับกระทรวงวิทยาศาสตร์ฯ กำลังจัดเก็บลงในสื่อคอมพิวเตอร์ อิเล็กทรอนิกส์ดิกชันนารีประเภทสำหรับมนุษย์ใช้นี้ ถ้าเก็บแบบภาษาเดียวก็จะเรียกว่า โมโนลิงกัวดิกชันนารี ถ้าเก็บแบบการแปลคู่ภาษาก็เรียกว่า ไบลิงกัวดิกชันนารี เช่น พจนานุกรมไทย-อังกฤษ การเก็บ

แสดงตัวอย่างหน้าจอภาพของระบบแปลภาษาด้วยเครื่องระบบ PENSÉE ของโอกิ



งานแสดงนิทรรศการส่วนใหญ่เป็นการแสดงผลงานทางด้านระบบการแปลภาษาด้วยเครื่อง



ใส่คอมพิวเตอร์ก็เพื่อช่วยให้ค้นหาและใช้งานได้สะดวกรวดเร็ว อิเล็กทรอนิกส์ดิกชันนารีนี้รวมไปถึงการใช้ในการตรวจสอบตัวสะกด (spell check) พจนานุกรมคำพ้อง (thesaurus) นอกจากนี้ยังเลยไปถึงการจัดเก็บข้อมูลจำนวนมากที่ใช้ในสารานุกรม ปัจจุบันมีการขายพจนานุกรมในรูปของสื่ออิเล็กทรอนิกส์แบบต่าง ๆ เช่น CD-ROM, เทป, ดิสค์ รวมไปถึงการสร้างพจนานุกรมฉบับเล็ก ๆ ที่เป็นคล้ายเครื่องคิดเลข

อิเล็กทรอนิกส์ดิกชันนารีอีกประเภทหนึ่งคือดิกชันนารีสำหรับเครื่องจักรใช้ เป็นพจนานุกรมที่ญี่ปุ่นกำลังให้ความสนใจสูงมาก ถึงกับลงทุนด้วยเงินหลายร้อยล้านบาทตั้งสถาบัน EDR ขึ้นมาเพื่อจัดทำอิเล็กทรอนิกส์ดิกชันนารีขายให้บริษัทต่าง ๆ ไว้ใช้ในงานประมวลผลข้อมูลภาษาธรรมชาติ ดิกชันนารีนี้จะเก็บข้อมูลเกี่ยวกับคำ ระดับไวยากรณ์ ระดับความหมาย และมุ่งลงไปถึง concept มีการแทนโครงสร้างข้อมูลที่ซับซ้อนเพื่อเป็นฐานความรู้สำหรับการประมวลผลภาษาธรรมชาติต่อไป ปัจจุบัน EDR ได้ประกาศว่าตนสามารถเก็บคำภาษาญี่ปุ่น-อังกฤษไว้แล้วประมาณ 200,000 คำ และเก็บ concept ไว้ได้ประมาณ 400,000 concept โครงสร้างของข้อมูลแต่ละระดับจะซอยกันอย่างละเอียด จำเป็นต้องอาศัยความรู้ความสามารถและแรงงานมหาศาล ซึ่งทางประเทศญี่ปุ่นเข้าใจดีจึงต้องลงขันกันให้ EDR เป็นผู้ดำเนินการให้

ระหว่างงานประชุมสัมมนาครั้งนี้ มีบริษัทผู้ผลิตคอมพิวเตอร์ของญี่ปุ่นประมาณ 10 บริษัทเข้าร่วมแสดงผลงานในบูตของตน ผลงานที่นำมาแสดงส่วนใหญ่เป็นเรื่องของการแปลภาษา สิ่งที่น่าสนใจมีหลายระบบ เช่น ของฟูจิตสึใช้ระบบไมโครต่อกันเป็น LAN ให้แต่ละสถานีแทนภาษาแต่ละภาษา เมื่อป้อนข้อมูลภาษาใดไปก็จะถูกแปลให้กับสถานีอื่นได้ทันที การแปลที่เห็นทั่วไปอยู่ในระดับจำกัด แต่ก็เห็นได้ว่าทางญี่ปุ่นมีความก้าวหน้าในด้านงานแปลภาษาไปในระดับหนึ่ง

จุดที่น่าสนใจของงานอีกจุดหนึ่งคือ EDR ได้นำเครื่องคอมพิวเตอร์ที่เก็บดิกชันนารีมาให้ผู้เข้าชมเล่นอักษรไขว้ปริศนา โดยเครื่องจะสร้างตารางแมทริกซ์ตามที่เราต้องการ จากนั้นก็ให้ผู้เล่นระบายสีดำลงในช่องที่ไม่ต้องการ คอมพิวเตอร์จะสร้างคำใส่ช่องไว้ภายในและจะบอกเราด้วยคำอธิบายทางขวางและทางดิ่ง โดยใช้ข้อความคำอธิบายของ concept ของคำ นับว่าเป็นการประยุกต์อิเล็กทรอนิกส์ดิกชันนารีให้เล่นกันได้ดีทีเดียว

ผลจากการประชุมนี้ทำให้ผู้เขียนเห็นว่าภาษาไทยเป็นภาษาประจำชาติ ความรู้ทางภาษาไทยนั้นเราคนไทยรู้ดีที่สุด การใช้เทคโนโลยีมาช่วยในการประมวลผลภาษาไทยนั้นนับวันจะมีความจำเป็นยิ่ง ๆ ขึ้นไป ดังนั้นเราควรจะมาสนใจและจัดทำอิเล็กทรอนิกส์ดิกชันนารีในภาษาไทยเพื่อใช้ประโยชน์และหาทางประยุกต์ใช้งานให้กว้างขวาง ผู้เขียนหวังว่าบทความเล่าสู่กันฟังครั้งนี้น่าจะกระตุ้นให้เราคนไทยหันมาสนใจภาษาไทยในลักษณะการประมวลผล สำหรับรายละเอียดทางด้านการแปลภาษาด้วยเครื่องและโครงสร้างอิเล็กทรอนิกส์ดิกชันนารีนั้นผู้เขียนจะนำมาเสนอในโอกาสต่อไป